УДК 519.218

ЗАБОЛОТНЫЙ С.В., к.т.н., доцент
САЛЫПА С.В., ассистент
ПЛАКСЕНКОВ Ю.Ю., магистрант (Черкасский государственный технологический университет)

Полиномиальные алгоритмы апостериорного оценивания момента разладки дисперсии негауссовских случайных последовательностей

Применен метод максимизации полинома для синтеза адаптивных алгоритмов апостериорного оценивания момента разладки дисперсии случайных последовательностей. Результаты статистического моделирования свидетельствуют о существенном повышении точности полиномиальных оценок, которое достигается путем учета негауссовости статистических данных.
Ключевые слова: *разладка, стохастический полином, дисперсия, кумулянтные коэффициенты.*

Вступление

Методы обнаружения и определения момента резкого изменения (разладки) стохастических сигналов имеют самые разнообразные области применения: радиолокация и радиоконтроль, обработка информации в автоматизированных системах управления, контроль качества технологических процессов, обнаружение аварийных ситуаций, текущий контроль систем подвижных объектов, обнаружение сейсмических сигналов на фоне помех, неразрушающий контроль, экологический мониторинг, медицинская диагностика и эконометрия [1 - 4].

Широкий спектр задач требует разработки большого количества разнообразных математических моделей и аппаратов их статистической обработки. При этом необходимо отметить, что подавляющее большинство фундаментальных теоретических работ, посвященных решению задач, связанных с «разладкой», сосредоточено на классе процессов, который описывается нормальным (гауссовским) законом распределения вероятностей. Однако значительная часть реальных процессов в информационно-коммуникационных каналах, в частности, на железнодорожном транспорте, отличаются от гауссовской модели [5]. Типичным примером является задача последовательного обнаружения сигналов контроля состояния рельсовой линии автоматической системы автоблокировки в условиях действия негауссовских помех, рассмотренная в работе [6]. Большинство методов, которые используют при решении задачи в подобных ситуациях, базируются на аппарате плотности или функции распределения вероятностей, что часто приводит к существенной сложности их практической реализации. Другой проблемой, характерной для параметрических методов (байесовского подхода, максимальной правдоподобности и др.), является необходимость наличия априорной информации о виде законов распределения и значениях их параметров.

Поэтому значительная часть современных исследований связаны с построением статистических методов, которые дают возможность убрать или минимизировать требования относительно объема априорной информации. Такие методы базируются на робастных процедурах статистической обработки, устойчивых к «неточности» вероятностных моделей, либо на непараметрических критериях, свободных от привязки к конкретным типам распределений. Очевидной платой за «не учет» вероятностных свойств статистических данных является ухудшение качественных характеристик сравнительно с оптимальными параметрическими методами.

Одним из компромиссных подходов к решению задач, связанных с обработкой негауссовских сигналов, является направление, базирующееся на применении аппарата стохастических полиномов Кунченка и моментно-кумулянтного описания [7]. Поскольку такое описание не является полным, то существует лишь асимптотическая (с ростом количества параметров) возможность получения оптимальных результатов. Однако использование данного подхода значительно упрощает процесс синтеза адаптивных статистических алгоритмов, способных на основе учета вероятностных свойств данных улучшать точность обработки (уменьшать вероятность ошибочных решений и дисперсию оценок).

В данной работе рассматривается апостериорный (ретроспективный) вариант постановки задачи о разладке, характерной особенностью которого является наличие выборки фиксированного объема, которая содержит сегменты статистических данных, отличающиеся определенными вероятностными характеристиками. Основным критерием анализа является точность (минимум дисперсии оценок) определения момента резкого изменения этих характеристик [1].

Целью данной работы является синтез на основе метода максимизации полинома алгоритмов апостериорного оценивания момента разладки по дисперсии случайных негауссовских последовательностей, а также исследования эффективности их применения путем статистического моделирования.

Математическая постановка задачи

Пусть наблюдается случайная последовательность (выборка) $\vec{x} = \{x_1, x_2, \ldots x_n\}$. Элементы этой выборки можно интерпретировать как совокупность n независимых случайных величин, вероятностный характер которых описывается математической ожиданием θ, дисперсией σ^2 и кумулянтными коэффициентами γ_l до определенного порядка $l = \overrightarrow{3,2s}$. До некоторого (априорно неизвестного) момента дискретного времени τ величина дисперсии равна σ_0^2, а затем, в момент $\tau + 1$ резко изменяет свое значение до величины σ_1^2. Необходимо на основе анализа всей выборки $\vec{x}$ оценить момент разладки τ.

Отдельные варианты апостериорной постановки задачи о разладке могут отличаться наличием априорной информации о значениях вариативного параметра (до и/или после разладки), а также о вероятностном характере других параметров модели случайной последовательности.

В данной работе рассматривается ситуация, когда значения всех параметров модели (кроме дисперсии) являются априорно известными и неизменными во времени.

Апостериорное оценивание момента разладки дисперсии методом максимального правдоподобия

Одним из основных направлений в исследовании апостериорных задач о разладке стал подход, который базируется на идее максимизации правдоподобия. Он был детально разработан Хинкли [8], который предложил общий асимптотический способ получения распределений апостериорных оценок момента разладки методом максимального правдоподобия (ММП). Применение данного подхода требует априорной информации о законе распределения статистических данных до и после разладки.

Известно, что при гауссовском распределении оценка дисперсии ММП при известном значении математического ожидания θ имеет вид

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{v=1}^{n} \left(x_v - \theta \right)^2 . \qquad (1)$$

Оценка вида (1) является состоятельной и асимптотически не смещенной, и как непараметрическая оценка метода моментов (ММ) может быть применена для оценивания дисперсии случайных величин с произвольным распределением. Однако эффективной такая оценка является лишь в случае гауссовской модели [9]. Для этой вероятностной модели логарифм функции максимального правдоподобия записывается в виде

$$\ln L\left(\sigma_0^2, \sigma_1^2\right) = -\frac{n}{2} \ln(2\pi) - \frac{r}{2} \ln\left(\sigma_0^2\right) - \frac{n-r}{2} \ln\left(\sigma_1^2\right) - \frac{\sum_{v=1}^{r} \left(x_v - \theta\right)^2}{2\sigma_0^2} - \frac{\sum_{v=r+1}^{n} \left(x_v - \theta\right)^2}{2\sigma_1^2} .$$

Для ситуации, когда величина θ является априорно известной, выборку $\vec{x}$ можно отцентрировать, обеспечив ее среднее значение равным нулю. При этом можно сформировать статистику вида

$$T_r\left(\sigma_0^2, \sigma_1^2\right) = -n \ln(2\pi) - r \ln\left(\sigma_0^2\right) - (n-r) \ln\left(\sigma_1^2\right) - \frac{\sum_{v=1}^{r} x_v^2}{\sigma_0^2} - \frac{\sum_{v=r+1}^{n} x_v^2}{\sigma_1^2} , \qquad (2)$$

которая будет принимать максимум в окрестности истинного значения момента разладки τ.

При априорно неизвестных значениях дисперсии случайной последовательности до и после разладки, в алгоритме могут быть использованы их текущие (апостериорные) оценочные значения этих параметров

$$\hat{\sigma}_{0,r}^2 = \frac{1}{r} \sum_{v=1}^{r} x_v^2 \qquad \hat{\sigma}_{1,r}^2 = \frac{1}{n-r} \sum_{v=r+1}^{n} x_v^2 . \qquad (3)$$

Результирующий алгоритм оценивания момента разладки будет иметь следующий вид [10]:

$$\hat{\tau} = \arg\max_{1\le r\le n-1} \left[n\ln\left(\hat{\sigma}^2\right) - r\ln\left(\hat{\sigma}^2_{0,r}\right) - (n-r)\ln\left(\hat{\sigma}^2_{1,r}\right)\right], \qquad (4)$$

где оценочное значение $\hat{\sigma}^2$ вычисляется для всей выборки по формуле (1).

Поскольку статистики в (2) и (4) не зависят от каких-либо иных параметров, то данные результаты могут быть применены для непараметрического оценивания момента разладки по дисперсии для случайных последовательностей с произвольным распределением. Однако, в такой ситуации, как и при оценивании дисперсии на основе (1), непараметрические алгоритмы в целом, теряют свою оптимальность, т.е. могут обеспечить точность, которая будет значительно меньше точности эффективных оценок. Потому ниже рассматриваются новые алгоритмы оценивания, базирующиеся на основе метода максимизации полинома (ММПл), применение которых позволяет достаточно просто учитывать степень негауссовости статистических данных [11].

Общий алгоритм полиномиального оценивания момента разладки

В соответствии с ММПл [7], при одинаково распределенных элементах выборки $\vec{x}$ оценка произвольного параметра ϑ может быть найдена из решения стохастического уравнения относительно оцениваемого параметра вида

$$\sum_{i=1}^{s} h_i(\vartheta)\left[\frac{1}{n}\sum_{v=1}^{n} x_v^i - \alpha_i(\vartheta)\right]\Bigg|_{\vartheta_r=\hat{\vartheta}_r} = 0,$$

где s – степень полинома, применяемого для оценки параметра;

$\alpha_i(\vartheta)$ – теоретические начальные моменты i-го порядка.

Коэффициенты $h_i(\vartheta)$, $i=\overline{1,s}$ находятся из решения системы ЛАУ, формируемой из условия обеспечения минимума (при соответствующей степени s) дисперсии оценки искомого параметра ϑ, вида

$$\sum_{i=1}^{s} h_i(\vartheta)\, F_{i,j}(\vartheta) = \frac{d}{d\vartheta}\alpha_j(\vartheta),\ j=\overline{1,s}, \qquad (5)$$

где $F_{i,j}(\vartheta) = \alpha_{i+j}(\vartheta) - \alpha_i(\vartheta)\alpha_j(\vartheta)$ – центрованные корелянты размерности (i,j).

Аналитическое решение системы уравнений (5) может быть найдено по методу Крамера

$$h_i(\vartheta) = \frac{\Delta_{is}}{\Delta_s},\ i=\overline{1,s},$$

где $\Delta_s = \det\left\|F_{i,j}\right\|$ $i,j=\overline{1,s}$ – объем тела стохастического полинома размерности s,

Δ_{is} – детерминант, который получается из Δ_s заменой i-го столбца на столбец свободных членов системы уравнений (5).

В данной работе предлагается новый подход для нахождения апостериорных оценок момента разладки, базирующийся на применении ММПл. Этот подход использует свойство стохастических полиномов вида

$$l_{sn}(\vec{x}/\vartheta) = nk_0(\vartheta) + \sum_{i=1}^{s} k_i(\vartheta)\sum_{v=1}^{n} x_v^i, \qquad (6)$$

где

$$k_0(\vartheta) = \int_a^{\vartheta}\sum_{i=1}^{s}\left[h_i(\vartheta)\alpha_i(\vartheta)\right]d\vartheta, \quad k_i(\vartheta) = \int_a^{\vartheta} h_i(\vartheta)\,d\vartheta$$
$$i=\overline{1,s}, \qquad (7)$$

которое заключается в том, что их математическое ожидание $E\{l_{sn}(\vec{x}/\vartheta)\}$ как функция ϑ принимает максимум в точке истинного значения этого параметра.

Если в качестве параметра ϑ, истинное значение которого, принадлежащее некоторому интервалу (a,b), будет максимизировать стохастический полином вида (6), использовать параметр, по которому происходит разладка (резкое изменение значения параметра с ϑ_0 на ϑ_1), то можно построить полиномиальную статистику вида

$$P_r^{(s)}(\vartheta_0,\vartheta_1) = rk_0(\vartheta_0) + \sum_{i=1}^{s} k_i(\vartheta_0)\sum_{v=1}^{r} x_v^i + (n-r)k_0(\vartheta_1) + \sum_{i=1}^{s} k_i(\vartheta_1)\sum_{v=r+1}^{n} x_v^i, \qquad (8)$$

которая будет иметь максимум в окрестности истинного значения момента разладки τ.

Таким образом, общий алгоритм нахождения оценки момента разладки на основе применения ММПл может быть формализирован в виде

$$\hat{\tau} = \arg\max_{1\le r\le n-1} P_r^{(s)}(\vartheta_0,\vartheta_1). \qquad (9)$$

Полиномиальное оценивание момента разладки дисперсии при $s = 2$

Известно [7, 11], что оценка дисперсии σ^2 ММПл может быть найдена лишь для степеней $s \geq 2$. При $s = 2$ полиномиальная оценка данного параметра (при $\theta = 0$) записывается в виде

$$\hat{\sigma}^2_{s=2} = \frac{1}{n}\sum_{v=1}^{n} x_v^2 + \frac{1}{2}\gamma_3^2\left(\frac{1}{n}\sum_{v=1}^{n} x_v\right)^2 - \\ -\gamma_3\left(\frac{1}{n}\sum_{v=1}^{n} x_v\right)\left\{\frac{1}{4}\gamma_3^2\left(\frac{1}{n}\sum_{v=1}^{n} x_v\right)^2 + \frac{1}{n}\sum_{v=1}^{n} x_v^2\right\}^{\frac{1}{2}}. \quad (10)$$

Анализ (10) показывает, что единственным дополнительным параметром, который влияет на нахождение оценки $\hat{\sigma}^2_{s=2}$, является коэффициент асимметрии γ_3. Тождественность значения этого параметра нулю (что соответствует симметрии распределения) приводит к вырождению полиномиальной оценки (10) в классическую оценку вида (1).

Показано [11], что использование (10) обеспечивает рост точности (уменьшение дисперсии оценки) сравнительно с оценкой (1), величина которой в асимптотическом случае (при $n \to \infty$) определяется теоретической зависимостью вида

$$g_2(\gamma_3, \gamma_4) = 1 - \frac{\gamma_3^2}{2 + \gamma_4}.$$

На основе аналитических выражений (5) и (7) можно легко найти, что коэффициенты, которые при степени $s = 2$ максимизируют выборочный стохастический полином вида (6) в окрестности истинного значения параметра σ^2, имеют следующий вид:

$$k_0(\sigma) = \frac{-\ln(\sigma^2)}{(2 + \gamma_4 - \gamma_3^2)}, \qquad k_1(\sigma) = \frac{2\gamma_3}{\sigma(2 + \gamma_4 - \gamma_3^2)},$$

$$k_2(\sigma) = \frac{-1}{\sigma^2(2 + \gamma_4 - \gamma_3^2)}.$$

Поскольку выражение $(2 + \gamma_4 - \gamma_3^2)$, не зависящее от σ^2, будет присутствовать в знаменателе каждой составляющей (8), то при построении статистик им можно пренебречь. Таким образом, при наличии априорной информации о значении дисперсии до σ_0^2 и после σ_1^2 разладки, полиномиальную статистику (8) при степени $s = 2$ можно записать в виде

$$P_r^{(2)}(\sigma_0^2, \sigma_1^2) = -r\ln(\sigma_0^2) + 2\gamma_3\frac{\sum_{v=1}^{r} x_v}{\sigma_0} - \frac{\sum_{v=1}^{r} x_v^2}{\sigma_0^2} - \\ -(n-r)\ln(\sigma_1^2) + 2\gamma_3\frac{\sum_{v=r+1}^{n} x_v}{\sigma_1} - \frac{\sum_{v=r+1}^{n} x_v^2}{\sigma_1^2}. \quad (11)$$

Можно показать, что при априорно известных значениях величин дисперсии случайной последовательности до и после разладки, эквивалентной (11), но более удобной в вычислительном плане, является статистика вида

$$R_r^{(2)}(\sigma_0^2, \sigma_1^2) = -(n-r)\left[\ln(\sigma_1^2) - \ln(\sigma_0^2)\right] + \\ + 2\gamma_3\left[\frac{1}{\sigma_1} - \frac{1}{\sigma_0}\right]\sum_{v=r+1}^{n} x_v - \left[\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right]\sum_{v=r+1}^{n} x_v^2. \quad (12)$$

При априорно неизвестных значениях дисперсии случайной последовательности, полиномиальная оценка момента разладки (как и в классическом случае) может быть найдена путем замены неизвестных величин дисперсии на их апостериорные оценки вида (3), которые формируются для каждого потенциального момента разладки.

Если априорно неизвестным является лишь значение дисперсии после разладки, то алгоритм нахождения оценки момента разладки может быть представлен в виде

$$\hat{\tau} = \arg\max_{1 \leq r \leq n-1} -\left\{(n-r)\left(\ln(\hat{\sigma}^2_{1,r}) - \ln(\sigma_0^2) + 1 - \frac{\hat{\sigma}^2_{1,r}}{\sigma_0^2}\right) - 2\gamma_3\left(\frac{1}{\hat{\sigma}_{1,r}} - \frac{1}{\sigma_0}\right)\sum_{v=r+1}^{n} x_v\right\}. \quad (13)$$

При априорно неизвестных значениях параметров до и после разладки алгоритм оценивания должен быть записан в следующей форме:

$$\hat{\tau} = \arg\max_{1\le r\le n-1}\left[2\gamma_3\left(\frac{\sum_{v=1}^{r}x_v}{\hat{\sigma}_{0,r}}+\frac{\sum_{v=r+1}^{n}x_v}{\hat{\sigma}_{1,r}}\right)-r\ln\left(\hat{\sigma}_{0,r}^2\right)-(n-r)\ln\left(\hat{\sigma}_{1,r}^2\right)-n\right]. \qquad (14)$$

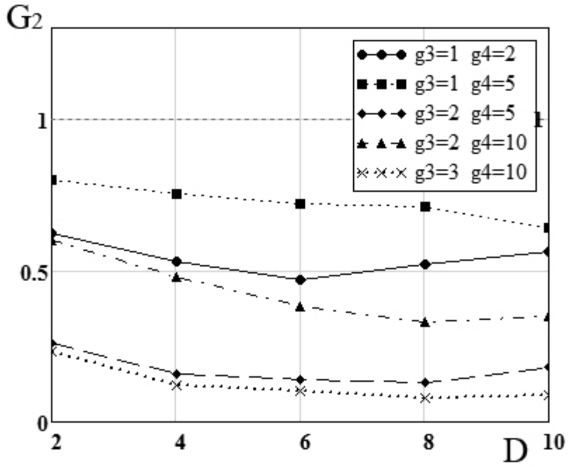
Сравнительный анализ структуры полиномиальных статистик (11) - (14) с классическими статистиками (2) и (4) еще раз подтверждает тот факт, что применение ММПл при степени полинома $s=2$ является целесообразным лишь при асимметрии ($\gamma_3 \ne 0$) распределения статистических данных.

Статистическое моделирование апостериорного оценивания момента разладки

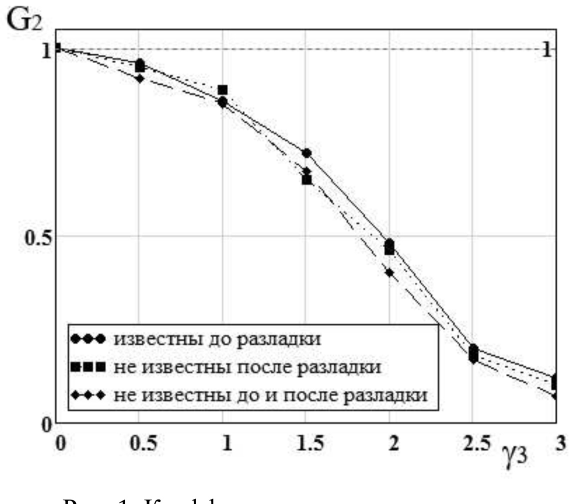
На основе результатов проведенных исследований, в программной среде MATLAB, был разработан комплекс имитационного статистического моделирования, который путем многоразовых экспериментальных испытаний методом Монте-Карло дает возможность осуществлять анализ точности предложенных полиномиальных алгоритмов апостериорного оценивания момента разладки. Критерием эффективности таких алгоритмов является отношение дисперсий полиномиальных оценок момента разладки $\hat{\tau}$ к дисперсиям классических оценок максимального правдоподобия, сформированных при условии гауссовского распределения модели.

Необходимо отметить, что на результаты статистического моделирования алгоритмов апостериорного оценивания момента разладки теоретически может влиять большое количество факторов, среди которых: степень (величина) разладки, вероятностный характер (значение кумулянтных коэффициентов высших порядков) негауссовских случайных последовательностей, наличие априорной информации о значениях вариативных параметров. Кроме того, на точность оценивания момента разладки также влияет общий объем выборки n, а на точность определения дисперсий оценок – количество экспериментов m, проводимых при одинаковых начальных условиях.

В качестве примера результатов статистического моделирования, на рис. 1 приведены экспериментальные значения (полученные при $n=200$ и $m=2000$) величин G_2, показывающих уменьшение дисперсии оценок момента разладки, которое обеспечивает применение ММПл при степени полинома $s=2$.



а)



б)

Рис. 1. Коэффициент уменьшения дисперсии оценок момента разладки

На рис.1, *а* приведены зависимости G_2 от относительной величины разладки $D=\sigma_1^2/\sigma_0^2$, полученные при различных значениях коэффициентов асимметрии γ_3 и эксцесса γ_4, а на рис.1, *б* – зависимости G_2 от γ_3 (при $\gamma_4=10$ и $D=4$), полученные при условиях различного объема информации о значениях дисперсии случайных последовательностей до и после разладки.

Анализ приведенных и других полученных экспериментальных результатов подтверждает теоретическое предположение об эффективности полиномиальных оценок момента разладки. При этом относительный рост точности является приблизительно одинаковым для разных постановок задачи, связанных с наличием/отсутствием априорной информации о значениях изменяемого параметра. Величина выигрыша существенно не зависит от относительной величины разладки, а определяется, прежде всего, степенью негауссовости, численно выражаемой абсолютными значениями кумулянтных коэффициентов.

Выводы

В целом, совокупность результатов проведенных исследований позволяет сделать общий вывод о потенциально высокой эффективности применения метода максимизации полинома для синтеза реализационно-простых адаптивных алгоритмов оценивания моментов разладки параметров случайных процессов, ориентированных на работу в условиях негауссовского характера статистических данных.

Следующим этапом исследований должен стать анализ влияния точности определения параметров негауссовской модели (кумулянтных коэффициентов высших порядков) на устойчивость полиномиальных алгоритмов апостериорного оценивания моментов разладки.

Литература

1. Никифоров И. В. Последовательное обнаружение изменения свойств временных рядов / И. В. Никифоров. – М. : Наука, 1983. – 199 с.
2. Бассвиль М. Обнаружение изменения свойств сигналов в динамических системах / М. Бассвиль. – М. : Мир, 1989. – 278 с.
3. Жиглявский А. А., Красковский А. Е. Обнаружение разладки случайных процессов в задачах радиотехники / А. А. Жиглявский, А. Е. Красковский. – Л. : Изд-во Ленинградского университета, 1988. – 224 с.
4. Богатырев А. А., Филиппов Ю. Д. Стандартизация статистических методов управления качеством / А. А. Богатырев, Ю. Д. Филиппов. – М. : Изд-во Стандартов, 1989. – 120 с.
5. Теория передачи сигналов на железнодорожном транспорте : Учебное пособие для вузов ж.д. трансп. / Г. В. Горелов и др. – М. : Транспорт, 2001. – 415 с.
6. Беляков И. В. Обнаружение сигналов контроля состояния рельсовой линии в негауссовских помехах методом поиска разладки / И. В. Беляков // Транспорт Урала. – 2011. – № 4. – С. 26-28
7. Кунченко Ю. П. Стохастические полиномы / Ю. П. Кунченко. – К. : Наук. думка, 2006. – 275 с.
8. Hinkley D. Inference about the change-point in a sequence of random variables / D. Hinkley // Biometrika. 1970. V.57. No.1. P.1–17.
9. Крамер Г. Математические методы статистики / Г. Крамер. – М. : Мир, 1975. – 648 с.
10. Chen J., Gupta A. K. Parametric statistical change point analysis / J. Chen, A. K. Gupta. – Birkhaeuser, 2000, 184 p.
11. Полиномиальные оценки параметров близких к гауссовским случайных величин. Ч II. // Ю. П. Кунченко, С. В. Заболотный. Оценка параметров близких к гауссовским случайных величин. – Черкассы : ЧИТИ, 2001. – 251 с.

Заболотній С.В., Салипа С.В., Плаксенков Ю.Ю. Поліноміальні алгоритми апостеріорного оцінювання моменту розладки дисперсії негаусових випадкових послідовностей. Застосовано метод максимізації полінома для синтезу адаптивних алгоритмів апостеріорного оцінювання моменту розладки дисперсії випадкових послідовностей. Результати статистичного моделювання свідчать про суттєве підвищення точності поліноміальних оцінок, яке базується на урахуванні негаусовості статистичних даних.

Ключові слова: розладка, стохастичний поліном, дисперсія, кумулянтні коефіцієнти.

Zabolotnii S.V., Salypa S.V., Plaksenkov Y.Y. Polynomial algorithms for a posteriori estimation of the moment of dispersion dissonance of Non-Gaussian random sequences. The Method of Polynom Maximization for the synthesis of adaptive algorithms of a posteriori estimation of the moment of dispersion dissonance of random sequences has been applied. The results of statistical modelling showed a significant increase of the polynomial estimators precision, which is reached by means of the record of Non-Gaussian statistic data.

Key words: dissonance, stochastic polynom, dispersion, cumulant coefficients.

Рецензент д.т.н., професор, завідувач кафедри СКС Лукашенко В.М. (ЧДТУ)

Поступила 24.10.2013г.